УДК 004.056.5

С. К. Алабугин
Научный руководитель: канд. тех. наук, доц. А. Н. Соколов
Южно-Уральский государственный университет, Челябинск

ПРИМЕНЕНИЕ МЕТОДОВ ОДНОКЛАССОВОЙ КЛАССИФИКАЦИИ В ЗАДАЧАХ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ

Аннотация. Работа посвящена использованию алгоритмов классификации в задачах обнаружения вторжений. Описаны проблемы, возникающие при использовании алгоритмов классификации в задачах обнаружения вторжений. В качестве метода решения предложено применение алгоритмов одноклассовой классификации. Приведены примеры таких алгоритмов, сформулирован и описан оригинальный подход к применению методов одноклассовой классификации в задачах обнаружения вторжений.

Ключевые слова: обнаружение вторжений; выявление аномалий машинное обучение; одноклассовая классификация; анализ сетевого трафика; обучение с учителем.

Вместе с ростом роли компьютерных сетей в жизни современного общества увеличивается и число сетевых атак, направленных на нарушение целостности, конфиденциальности и доступности информации. При этом появляются новые, более комплексные, виды атак, в частности получили распространение целевые атаки (Advanced Persistent Threat).

Поэтому на сегодняшний день задачи обнаружения вторжений (в частности, сетевых атак) являются крайне актуальными. Классические методы, применяемые для анализа сетевого трафика, которые основаны на использовании сигнатур атак, не в состоянии обнаружить новую, до этого момента не известную атаку. В связи с этим большое количество исследований посвящено методам обнаружения вторжений через выявление аномалий сетевого трафика, в частности с помощью техник и методов машинного обучения. Стоит отметить, что данный подход обладает рядом недостатков по сравнению с использованием сигнатур атак (большее число ложных тревог, большие затраты вычислительных мощностей и др.).

Наибольшее число работ в данной области посвящены использованию классических алгоритмов классификации (метрические и линейные классификаторы, сети Байеса, деревья принятия решения, нейронные сети) для обнаружения атак в сетевом трафике. При этом используется традиционная методология ма-

шинного обучения: сетевой трафик (чаще всего это отдельное сетевое соединение) описывается вектором признаков — длительность соединения, IP-адреса хостов, количество переданных пакетов и т. п. Каждый вектор признаков, маркируется как безопасный или атака соответственно, затем на основе множества таких векторов (обучающая выборка) происходит обучение алгоритма. Однако при таком подходе приходится решать две основные проблемы.

Первая, более фундаментальная, заключается в том, что классические алгоритмы классификации по сути своей не приспособлены для нахождения новых атак. Так, алгоритмы, обученные на выборке, состоящей из объектов представляющих нормальный трафик и какие-то виды сетевых атак, вполне могут классифицировать объект, представляющий новую для него атаку как нормальный трафик.

Вторая, более практическая, происходит от того факта, что в каждой компьютерной сети существует свое понятие нормального трафика (это объясняется тем, что сети отличаются друг от друга топологией, используемыми техническими средствами и работающими в рамках сети приложениями). Таким образом, чтобы обучить алгоритм в рамках конкретной сети, нужен не только нормальный трафик, но и трафик, представляющий различные виды сетевых атак. Чтобы получить его, приходится либо самостоятельно проводить атаки в рамках своей сети, либо программно генерировать дампы сетевого трафика, который будет содержать реализации тех или иных атак.

Одним из подходов, который может использоваться для решения вышеописанных проблем, является одноклассовая классификация. В рамках этой техники обучение происходит на объектах лишь одного класса, а обученный алгоритм должен давать ответ, принадлежит ли новый объект этому классу или нет. Данная задача является более трудной, чем многоклассовая классификация, поэтому алгоритмов, решающих ее в рамках машинного обучения, существенно меньше. Для примера упомянем два алгоритма.

Модификация алгоритма *K* ближайших соседей основана на вычислении отношения расстояний между объектами в пространстве признаков. Пусть z — объект, который нужно классифицировать, y — его ближайший сосед (наиболее близкий объект согласно некоторой метрике, введенной в пространстве признаков) из обучающей выборки, $NN(y)$ — ближайший сосед y из обучающей выборки, d — метрика в пространстве признаков, тогда сравнение

$$\frac{d(z,y)}{d\big(y,NN(y)\big)} < \mu$$

определяет ответ алгоритма: в случае, если данное отношение меньше порога μ, то считается, что объект принадлежит классу.

Машина опорных векторов для одноклассовой классификации представляет обычную машину опорных векторов, однако гиперплоскость в пространстве признаков, которая строится алгоритмом, не разделяет объекты нескольких классов, но выделяет границы одного класса.

На начальном этапе исследований нами был сформулирован и реализован свой алгоритм одноклассовой классификации, основанный на оценке плотности распределения объектов и построении решающей границы класса методом выпуклой оболочки. Выпуклой оболочкой множества X называют наименьшее выпуклое множество, содержащее X. В частности, если X — множество точек на плоскости, то его выпуклой оболочкой будет замкнутая линия, проведенная через граничные элементы X. Основная идея предложенного нами подхода заключается в том, чтобы построить выпуклую оболочку в пространстве признаков относительно объектов, соответствующих нормальным соединениям. Классификация новых объектов осуществляется путем проверки вхождения их в эту замкнутую оболочку. Вообще говоря, выпуклая оболочка может быть построена в пространстве любой размерности, однако чем больше размерность, тем больше времени будет требоваться для построения оболочки и проверки вхождения в нее объекта. По этой причине, а также для большей наглядности было решено понизить размерность пространства признаков до двух с помощью метода главных компонент. Следующим шагом является построение выпуклой оболочки относительно нормальных объектов. Для того чтобы определить область наибольшего скопления этих объектов в пространстве признаков, была использована ядерная оценка плотности вероятности. Посчитав таким образом плотность вероятности для всех рассматриваемых объектов, мы можем определить те, что лежат в области наибольшей плотности и использовать их для построения выпуклой оболочки в первую очередь. В качестве настраиваемого параметра алгоритма мы вводим долю объектов выборки, которые будут использоваться для построения выпуклой оболочки. С увеличением параметра классификатор обнаруживает меньше атак и считает больше объектов нормальными. В ходе экспериментов на классическом наборе данных NSL KDD удалось добиться точности классификации 0.83. Для увеличения точности данного алгоритма стоит исследовать выбор признакового описания объектов, чтобы найти признаки, обеспечивающие лучшее разделение нормального сетевого трафика и аномалий.

Таким образом, применение алгоритмов одноклассовой классификации в задаче обнаружения вторжений позволяет обойти некоторые проблемы, возникающие при использовании многоклассовых классификаторов. Однако пока не существует алгоритмов, которые бы качественно выявляли атаки на основе анализа сетевого трафика.